Probabilistic Data-Driven Modelling of Astrophysical Transients: The Neural Process Family for Ultrafast and Class-Agnostic Light Curve Reconstruction

Siddharth Chaini [1] **Federica B. Bianco** [1 2] **Ashish Mahabal** [3]

Abstract

Astronomical light curves — the sparse, irregular, multi-wavelength time series produced by photometric surveys — are notoriously hard to model, with Gaussian Processes (GPs) the current standard despite their poor scaling, *a priori* kernel specification, and inability to share information across objects. We introduce the use of Attentive Neural Processes (ANPs) as a fast, data-driven, morphologically agnostic interpolator: a single meta-learned model handles 14 explosive transient classes without specifying the class at inference. Evaluated on realistic LSST cadences, the ANP outperforms seven GP- and neural-network-based benchmarks on all regression, feature-recovery, and probabilistic-calibration metrics, while running in $\sim 10^{-6}$ s per light curve — four to five orders of magnitude faster than benchmark models — fast enough for on-line use on the LSST alert stream of $\sim 10^7$ events/night. We demonstrate how the ANP avoids neural networks' overconfidence and GPs' underconfidence, delivering sharp, calibrated uncertainties suitable for downstream science. This work establishes the neural process family as a scalable, probabilistic foundation for real-time transient science in the LSST era.

1. Introduction and Background

The goal of time-domain astrophysics (TDA) is to understand how astronomical objects evolve over time. *Transients* — short lived brightening events such as supernovae and other explosive events — are a key focus; large astronomical surveys observe them in alternating photometric filters over time to probe their physics - like temperature or chemical composition, which is encoded in its spectral energy distribution (SED; energy as a function of wavelength and time). However, observations from ground-based observatories are subject to weather, lunar cycles, and scheduling constraints from other astronomy goals. This greatly impacts transient science; light curves — flux over time across multiple filters — are characteristically multivariate, sparse, and irregular, with heteroskedastic uncertainties from instrumental and environmental factors and the intrinsic properties of the source. This poses a great challenge for statistical and machine learning methods applied to light curves.

Just around the corner is the revolutionary Vera C. Rubin Observatory and its flagship astronomical survey, the Legacy Survey of Space and Time (LSST, Ivezić et al., 2019). LSST will observe the complete southern sky from Chile, revisiting each position every few nights for 10 years. For every change in brightness above 5 σ, it will generate an *alert* – we expect $\sim$10 M alerts per night, telling us about the different transients and variables in our Universe. LSST features six filters ($ugrizy$) spanning the wavelength range from the UV (320 nm) to the near IR (1,050 nm) .Its cadence, however, is subject to the constraints from other non time-domain science cases (Rubin's Survey Cadence Optimization Committee et al., 2025), and will be sparse even compared to many other time-domain surveys. Most light curves will have gaps comparable to the evolutionary time scales of the underlying physics, limiting what can be learnt from a single light curve. The sparse and irregularly sampled nature of the multi-filter LSST light curves limits the performance of statistical models on downstream tasks, including and perhaps especially machine learning approaches as light curves are poorly suited tensors for most neural network architectures. But what LSST lacks in cadence per object, it will make up for in number observing many millions of objects. Therefore, a method that can learn and carry information across different light curves is highly desirable.

Many traditional methods involve extracting features from parametric fits (*e.g.*, Malanchev et al., 2021; Van Roestel et al., 2021; Malanchev, 2021; Jainaga et al., 2021; Sánchez-Sáez et al., 2021; Healy et al., 2024; Chaini et al., 2024), but these are now insufficient - they are slow, do not work well for multiple filters, and fail for sparse data. Instead, the current standard is Gaussian Processes (GP, Rasmussen &

[1]University of Delaware [2]Rubin Observatory [3]Caltech. Correspondence to: Siddharth Chaini <chaini@udel.edu>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

Williams, 2008) – used for interpolation and reconstruction of light curves (Villar et al., 2021; Aigrain & Foreman-Mackey, 2023; Qu & Sako, 2022; Peng et al., 2024; Khakpash et al., 2024; Bhardwaj et al., 2025; Magill et al., 2026; Pellegrino et al., 2026). Yet, GPs require *a priori* definition of a kernel, and are computationally expensive since they are fit directly to each light curve, and highly susceptible to overfitting (*e.g.*, Blunt et al., 2023).

Deep learning has become a popular modern approach, with the allure of learning feature representations from the light curves automatically during training (*e.g.*, Mahabal et al., 2017; Muthukrishna et al., 2019; Möller & de Boissière, 2020; Shah et al., 2025; Donoso-Oliva et al., 2023; Cabrera-Vives et al., 2024; Donoso-Oliva et al., 2025), but more often than not comprise of bespoke models for a particular task. Meanwhile, general-purpose Time-Series Foundation Models (TSFMs) (*e.g.*, Das et al., 2023; Garza et al., 2023; Woo et al., 2024; Ansari et al., 2024) have been very successful at zero-shot performance across fields like finance and weather, but are difficult to fine-tune for light curves, struggling with sparsity and irregularity, as demonstrated by (Li et al., 2026). These deep learning approaches also bind the user to rely on neural networks, which, while increasingly explainable, make interpretation more difficult, as there is value in being able to 'inspect' the light curve to guide inference. Another key limitation of most deep learning approaches is the lack of reliable uncertainty estimates and, even when present, these uncertainties are known to suffer from miscalibration (Guo et al., 2017). For these reasons, the practice of interpolating, or better, 'reconstructing' light curves remains widely used as a preprocessing step before direct analysis and before applying machine and deep learning models.

To model class-agnostic light curves, emergent deep learning methods such as Bayesian neural nets and normalizing flows (Demianenko et al., 2023) have been proposed, alongside newer flexible generative models: diffusion-based interpolators that iteratively denoise a latent representation (Sasseville et al., 2026; Shen & Gagliano, 2025), neural-ODE-based methods that parameterise the flux evolution as a differential equation (Wu et al., 2026), and transformer-based interpolators that propagate uncertainties through attention layers (Yalavarthi et al., 2022; Sasseville et al., 2026) – yet the latter are shown to produce miscalibrated, overconfident uncertainties (Sasseville et al., 2026). Generally, the latency due to successive forward passes with these methods is also problematic when the goal is to apply them to large datasets in real time, such as the LSST alert stream.

Contribution. We propose the neural process family (Garnelo et al., 2018a;b; Kim et al., 2019) for multiwavelength, sparse and irregular light curve reconstruction. Neural Processes (NPs) are meta-learning models that combine the probabilistic flexibility of GPs with amortized neural inference, and are non-parametric and rapid at inference. They have seen limited use in the physical sciences – mostly for climate science (Andersson et al., 2023; Scholz et al., 2023), while prior use in astronomy is narrowly scoped to single-class settings (exoplanets (Gebhard et al., 2024), gravitational waves (Schuetz et al., 2025) and AGN/quasar variability (Park et al., 2021; Čvorović-Hajdinjak et al., 2022; Jankov et al., 2022; Kovacevic et al., 2023; Cvorovic-Hajdinjak, 2024). In these applications, NPs are trained on light curves from a single object class (*i.e.*, that share the same physical processes), and often used as an offline preprocessing step to simply extract parameters related to that class. **Here, we fundamentally shift this paradigm**, and propose ANPs as fast, performant and *universal* transient multifilter light-curve interpolators – a single, morphologically agnostic model that meta-learns a shared prior across a diverse population of explosive transients without prespecifying the class at inference. While we restrict this proof-of-concept to explosive transients, the data-driven framework is naturally extensible to broader classes of variable astronomical sources, and to downstream tasks of astrophysical relevance (parameter estimation, classification, anomaly detection) that we leave to future work.

This work is entirely reproducible; all data and code is available here[1], with architecture details given in App. B.

2. Data and Methods

Benchmarks. We compare ANPs with seven benchmark models (Sec. 3). We follow the comparison paradigm of (Demianenko et al., 2023): a Bayesian Neural Network (BNN), Normalising Flows (NF), two MLPs (PyTorch and scikit-learn), and a Gaussian Process (GP) with three kernel variants. We deliberately do not benchmark head-to-head against parametric-model interpolators such as Rainbow (Russeil et al., 2024), neural-ODE based class-conditioned methods (Wu et al., 2026), or recent diffusion-based interpolators (Shen & Gagliano, 2025; Sasseville et al., 2026). These methods address different problem formulations, rely on strict class-specific inductive biases, or operate under non-overlapping inference-time constraints: Rainbow is parametric, does not provide uncertainties and is class-specific. SELDON (Wu et al., 2026) is contemporaneous work that does not provide uncertainties, having been demonstrated on a single class so far; and the diffusion-based methods are orders of magnitude slower at inference, with (Sasseville et al., 2026) reporting that transformer-based variants produce miscalibrated, overconfident uncertainties.

[1] https://github.com/sidchaini/NightLANP

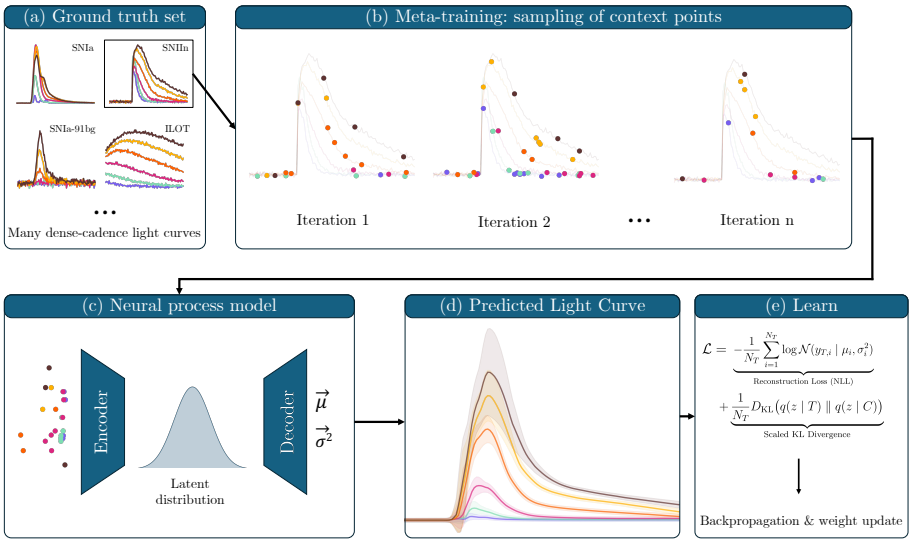

Figure 1. Schematic overview of the meta-learning training procedure for our Attentive Neural Process (ANP) model on light curves. (a) We start with a ground-truth set of 15,000 dense (1–2 day cadence) simulated light curves spanning 14 transient classes. (b) At every meta-training iteration, we randomly draw between 12 and 60 points from a slice of each light curve to form a context set. (c) The neural process model (Fig. 6) consists of an encoder-decoder architecture. (d) The decoder predicts a flux mean and variance at every queried time and filter. (e) The model is trained to minimise the negative log-likelihood ($\mathcal{L}_{\mathrm{NLL}}$) and KL divergence via backpropagation.

Related work: the Neural Process Family. The Neural Process family was introduced as Conditional Neural Processes (CNP, Garnelo et al., 2018a) and (Latent) Neural Processes (NP, Garnelo et al., 2018b), both of which combine the flexibility of GP-style probabilistic inference with the scalability of amortized neural inference. Attentive Neural Processes (ANP, Kim et al., 2019) replace mean-pooling aggregation with attention to address NP underfitting. Subsequent advances include convolutional CNPs (Gordon et al., 2020), autoregressive variants (Bruinsma et al., 2023), Gaussian NPs (Bruinsma et al., 2021), Transformer NPs (Nguyen & Grover, 2023), Rényi NPs (Wang et al., 2025), flow-matching NPs (Hamad & Rosenbaum, 2025), and incremental transformer NPs (Mortimer et al., 2026), among others. We adopt ANPs as a robust, well-understood baseline, and leave extensions to more recent variants to future work.

Traditional supervised models map fixed-size inputs to outputs. But for time-domain astrophysics, we need models mapping variable-length observations to a continuous function queryable at arbitrary time steps. GPs (Rasmussen & Williams, 2008) do this natively via kernels that model flux covariances across time and wavelength. Ideally, these kernels reflect the specific physics driving a transient (Foreman-Mackey et al., 2017; Khakpash et al., 2024). However, for modeling light curves in large astronomical surveys like LSST, an object's class is not known *a priori*, and so neither is the physics governing it. Moreover, different astronomical classes are generated by fundamentally different physical processes; thus no single physically motivated kernel can model an entire dataset. Hence GPs require fitting a poorly-constrained kernel to each light curve individually – the fundamental bottleneck for real-time processing of $\sim$10 M LSST alerts/night, and prevent the GP from learning shared representations across the survey.

NPs learn through *meta-learning* (Dubois et al., 2020): at every iteration (called an episode), a random context set is used to predict the corresponding targets along with an uncertainty (Fig. 1). An NP encoder maps context points $(\vec{x}_c, \vec{y}_c)_i$ to representations aggregated into a summary $\vec{r}_*$; a decoder takes $\vec{r}_*$ and query locations $\vec{x}_*$ to predict a mean and variance $(\vec{y}_\mu,\ \vec{y}_{\sigma^2})$. The Attentive Neural Process (ANP; Kim et al., 2019) uses self-attention, and replaces mean-aggregation with cross-attention (Vaswani et al., 2017), producing target-specific representations, and adds a latent path that captures epistemic uncertainty via a global latent variable $\vec{z}$ sampled from a learned posterior. The decoder concatenates query coordinates, the deterministic representation, and $\vec{z}$ to produce a predictive mean $\vec{\mathcal{F}}_\mu$ and aleatoric variance $\vec{\mathcal{F}}_{\sigma^2}$ (architecture diagram in App. B). We benchmarked CNP, NP, and ANP variants and found the ANP to converge to lower validation loss with stably calibrated uncertainties; we thus adopt the ANP for all subsequent experiments and leave other recent NP variants to future work.

Data. Our training data is a dataset of high-cadence simulated transient "light curves", or multivariate time series of light at different wavelengths (Ishida & Jones, 2025). These light curves have observations for all 6 LSST filters over

a 1–2 day cadence, derived from the PLAsTiCC simulation models (PLAsTiCC Modelers, 2022). An example of a dense-cadence light curve is shown by the crosses in Fig. 5.

The full dataset consists of nine different models including 14 astrophysical transient and variable phenomena (with the `NON1ASED.V19_CC+HostXT` model spanning 7 subclasses). The flux $y \in \mathbb{R}$ is treated as a deterministic input; the ANP does not currently ingest input flux uncertainties as additional features, instead *predicting* an output uncertainty through the decoder. In an effort to maximize the information contained in this subset, we selected the longest 1,250 light curves per class, which were then randomly split into 1,000 training and 250 testing light curves per class. Full details of our final dataset are given in Sec. A.1.

To evaluate our model, we use an unseen test set of light curves generated by degrading our dense-cadence test set into sparse, irregularly sampled light curves that mimic the observational cadence of the main LSST survey. We do this 10 times to generate 10 LSST-like *realizations* (between a minimum of 10 and a maximum of 141 points) for each test light curve. Full details are provided in Sec. A.2.

2.1. Adapting Neural Processes for Astronomical Time Series

Standard NP implementations in the ML literature are often highly abstracted to demonstrate GP emulation; to deploy them on real data we develop a package `keras-neural-processes`[2] (Keras 3 (Chollet et al., 2015), multi-backend across JAX (Bradbury et al., 2018), PyTorch (Paszke et al., 2019), and TensorFlow (Abadi et al., 2015), with ragged-tensor support), which is general and works with any input of arbitrary dimensions $\boldsymbol{x}$, to predict any output $\boldsymbol{y}$ of arbitrary dimensions, enabling broader use in the physical sciences. For our astronomical use-case, we have a separate wrapper, `NightLANP`[3], that implements the following four engineering adaptations for use with light curves.

Filter-as-coordinate. Our input is a 2D vector $\boldsymbol{x} = [t, b]$ consisting of the modified Julian date (MJD; t) and filter index $b \in \{u{:}0, g{:}1, r{:}2, i{:}3, z{:}4, y{:}5\}$, while the target $\boldsymbol{y}$ is the flux. We pass b as a raw integer scalar; and note the potential of using learned embeddings here, which if left for future work. The ANP nonetheless recovers cross-filter correlations directly from data (Sec. 3). Treating the filter as a coordinate (rather than a separate channel) elegantly handles ragged inputs and missing filters. While this would introduce redundancy if all six filters were observed simultaneously, such observations never occur in LSST-like cadences (at inference), meaning this theoretical cost is not realized in practice.

Dynamic slicing. At every episode we randomly crop each light curve to a continuous 678-point slice ($\sim$57 days; chosen empirically) – long enough to capture significant transient evolution but short enough to fit on two T4 GPUs. The slice forms the target set, which is randomly subsampled to draw the context points. This serves three purposes simultaneously: (i) full-length light curves are memory-prohibitive; (ii) fixed-shape target tensors enable efficient GPU parallelization despite irregular underlying sampling; and (iii) it prevents temporal bias where models cropped to peak-centred regions learn an implicit brightness relation, making them unusable on 'live' transients where the peak is unknown.

Temporal re-zeroing. While ANPs do not enforce translational equivariance implicitly, we implement a live augmentation where we subtract the slice start time before each batch. This forces the ANP to learn from relative times only, allowing for use on partial light curves without any time-alignment. We also note that because of this, the same light curve results in different target slices with different starting points over the course of training.

Stratified sampling. To handle class imbalance, we draw each batch with roughly equal class representation. While this may lead to over-representation from the minority classes, it ensures the majority classes do not dominate the loss. Combined with the randomization in dynamic slicing and context selection, the ANP effectively never sees the same temporal configuration twice .

Training Procedure. Our model performs dynamic slicing on the full dense-cadence training set; where the context points sampled within each slice are a small fraction (1.8%-8.8%) of the target points.[4] While forecasting peak time from rise-only observations is critical for astrophysical follow-up (Hambleton et al., 2022), it is unsurprisingly ill-suited when the context spans the full light curve. So, we additionally train a separate 'forecasting-mode' ANP where 50% of context points are pre-peak, and use that when forecasting predictions of the peak (Tab. 3.1).

Our ANP implementation uses a 4-layer MLP for the deterministic encoder and 2-layer MLPs for the latent encoder and decoder both, with 128 ReLU-activated hidden units throughout (output layers unactivated). The deterministic path uses self- and cross-attention with 8 heads; the latent path uses mean-pooling to parameterize a 128-dimensional global latent distribution.

[2] https://github.com/sidchaini/keras-neural-processes

[3] https://github.com/sidchaini/NightLANP

[4] We sample the context-set length from 10 fixed options rather than a continuous uniform distribution to avoid JIT recompilation on every new length.

We trained the ANP 700,000 episodes with batch size 64 on two NVIDIA T4 GPUs in 5 hours. Using the Adam optimizer at constant learning rate 10^{-4}, we minimize the negative log-likelihood ($\mathcal{L}_{\mathrm{NLL}}$) and KL divergence between the variational posterior and prior latent distributions. We did not perform an extensive architectural search; this is the first configuration that converged stably, and we expect further tuning would yield improvements. Nonetheless, the ANP outperforms all benchmark models (Demianenko et al., 2023) as demonstrated in Sec. 3.

3. Experiments & Results

The scientific objective is to reconstruct the complete dense multifilter light curve (the *target* set) from sparse LSST-cadence (*context points*) for all transients in the test set. We compare the ANP against seven benchmarks (see Sec. 2). An example reconstruction for each model is shown in Fig. 2.

3.1. Reconstruction Quality

We evaluate model predictions drawing from (Demianenko et al., 2023)'s evaluation procedure with slight modifications: for each of the 3,750 test objects we generate 10 independent LSST-cadence realizations to mitigate 'unlucky' samplings (*e.g.*, where context points are only sampled at the very beginning or at the tail end of the light curve). Treating these realizations as independent would inflate sample size due to pseudoreplication (Hurlbert, 1984); we therefore median-aggregate metrics over the 10 realizations before comparing across light curves for all models. All our aggregate metrics are reported in Tab. 1, and we report further details in App. C.

Standard regression metrics. Mean Squared Error (MSE), Mean Squared Scaled Error (MSSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE) evaluate point-prediction accuracy $\vec{\mathcal{F}}_\mu$ against ground truth, and are computed per object per filter then averaged across filters. Scaled variants (MSSE, MASE) are normalized by mean flux differences, making them robust to the natural variations in flux across transient classes. **The ANP achieves the lowest median and most compact distribution on all four**.

Uncertainty-aware regression metrics. χ^2, Normalized Root Mean Squared Error ($\mathrm{NRMSE_p}$ on the the model's predicted error, $\mathrm{NRMSE_o}$ on the simulation derived observed error) explicitly penalize predictions when the true flux falls outside the predicted error bounds. Being 'overconfidently wrong' wastes follow-up resources, so this matters in practice. **The ANP achieves the lowest median across all three**; BNN, MLP, and NF are orders of magnitude worse, reflecting catastrophic overconfidence (Guo et al., 2017), while GPs sit roughly $2\times$ above the ANP.

Peak-estimation metrics. This is a physically motivated set of metrics: $|\Delta t_{\mathrm{peak}}|$ and $|\Delta m_{\mathrm{peak}}|$ measure the error in time and brightness (luptitude scale with zero-point at 27.5; (Lupton et al., 1999)) of the predicted light curve peak. These are computed both globally (max flux across the six filters) and per filter (then averaged). The per-filter variant rewards correct color evolution. **ANP wins all four**.

Forecasting peak time from rise-only context. The astrophysically motivated case of forecasting peak time and brightness is critical for planning time-sensitive astronomical follow-up. When trained on context points throughout the light curve, ANP is unsurprisingly poor at forecasting; we therefore train a separate 'forecasting-mode' ANP where 50% of context points are pre-peak (Sec. 2.1). The results are reported in Fig. 12. **The forecasting ANP wins on both global and per-filter** (App. C). While GPs are more competitive on the global metric, they rapidly regress to baseline in the absence of observations; therefore, their accuracy is an artifact of how we constructed the context set. All models remain imperfect at forecasting; we leave improvements to future work.

3.2. Probabilistic Calibration

We follow (Gneiting et al., 2007)'s evaluation recommendation for probabilistic forecasts, using proper scoring rules which "maximize sharpness subject to calibration" (sharper error bars are preferred, but only if they contain the true value at the stated rate).

Coverage and width. The Prediction Interval Coverage Probability at 95% ($\mathrm{PICP}_{0.95}$ denotes the fraction of true values within the 95% interval) and Mean Prediction Interval Width (MPIW) denotes the uncertainty width. ANP is the best calibrated and closest to the ideal $\mathrm{PICP}_{0.95}{=}95\%$, while still showing a sharp MPIW. BNN, MLP, and NF have even narrower MPIWs, but very low PICP values – indicating they are confidently wrong.

Calibration curves. The empirical PICP is plotted in Fig. 3 (top) as a function of predicted confidence. ANP is closest to the ideal, with residuals (bottom) $< 5\%$ which decrease nearly monotonically with confidence, whereas BNN/MLPs are drastically overconfident; GP2/GP3 are next-best but oscillate at higher confidences.

Proper scoring rules. Negative Log Predictive Density (NLPD) and Continuous Ranked Probability Score (CRPS) are maximized only if the predicted probability distribution perfectly matches the target distribution (Gneiting &

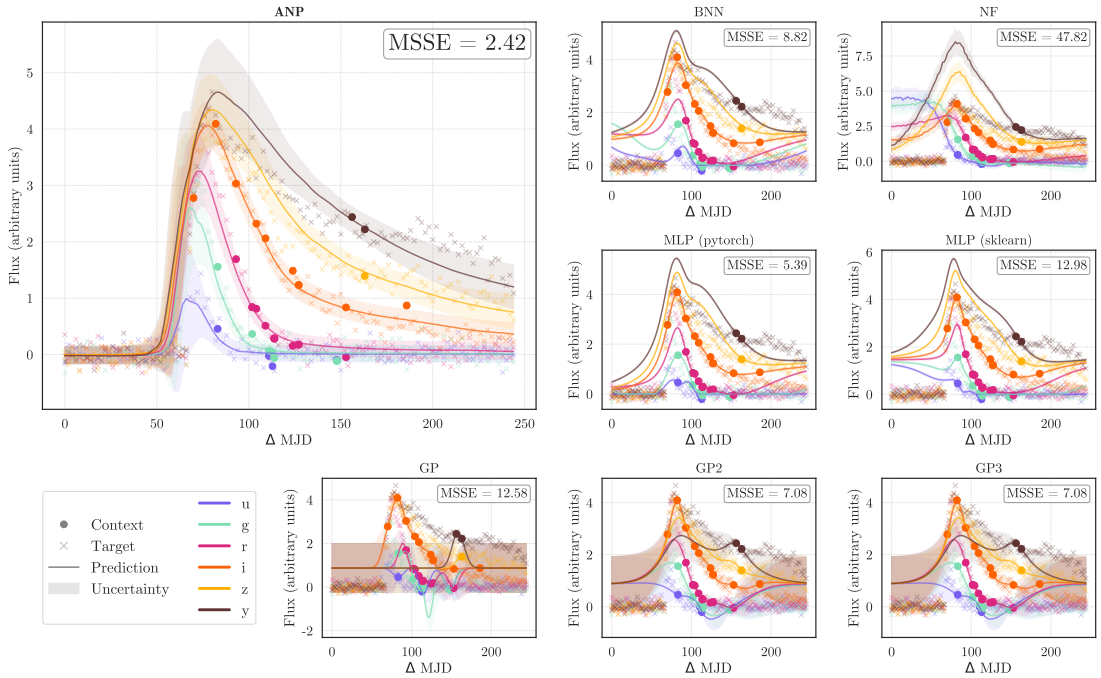

Figure 2. Model predictions for an example Type IIP supernova by the ANP and all seven comparison models considered in this work. The Mean Scaled Squared Error (MSSE) of each prediction is indicated. The predictions (means) are shown as solid lines, and $\pm 1\sigma$ uncertainty filters (shaded regions) are shown for each of the six LSST filters $ugrizy$, overlaid on the context points (circles) and target points (crosses). Our ANP result is shown in the top-left panel (enlarged). The remaining panels show the same light curve as interpolated by the seven benchmark models as implemented from (Demianenko et al., 2023) (see Sec. 3). The ANP outperforms all the models with the lowest MSSE (2.42), roughly capturing the rise and decay.

Raftery, 2007; Pion & Vazquez, 2025). NLPD penalizes predictions assigning low probability to true values, while CRPS generalizes MAE to distributions. ANP's NLPD is over an order of magnitude better than the next-best GP, and also wins on CRPS; ANP predictions are thus simultaneously sharp and centered on the truth.

PIT histograms. The Probability Integral Transform (Gneiting et al., 2007) – the predicted CDF evaluated at the ground-truth flux – is uniform on $[0, 1]$ for a perfectly calibrated model. Fig. 4 reports PIT histograms with the Kolmogorov–Smirnov (KS) statistic against Uniform$(0, 1)$.[5] BNN and MLPs show severe U-shapes (overconfidence boundary spikes); GPs are better (KS=0.134 for GP3) but left-skewed, indicating a positive flux-mean bias. The ANP achieves the flattest histogram (KS=0.048), avoiding both failure modes simultaneously.

3.3. Timing and Computational Complexity

The ANP, a meta-learning model unlike all the other benchmarks, has an extra meta-training phase. While this allows it

[5]None of these distributions ($> 10^6$ points) are statistically consistent with Uniform$(0, 1)$ at the $> 5\sigma$ level; we use KS as a similarity metric, not a consistency test.

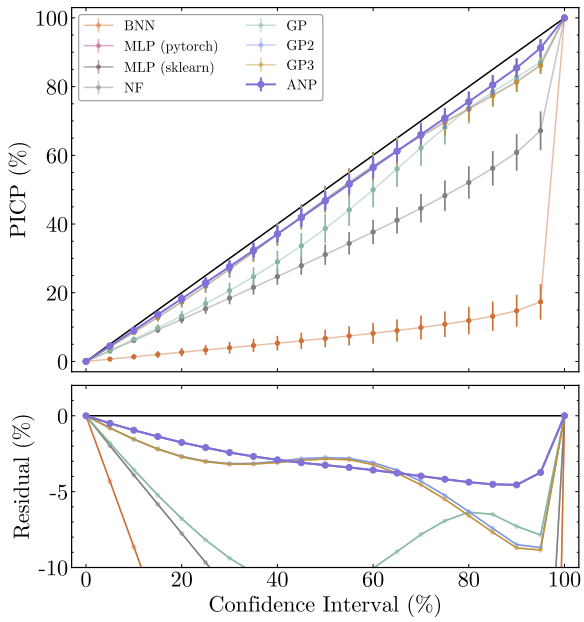

Figure 3. Empirical coverage PICP vs predicted confidence interval (top), with residuals (bottom). Median-aggregated over 10 realizations. ANP residuals are $< 5\%$ and nearly linear; BNN and MLPs are drastically overconfident; GP2/GP3 are next-best.

Table 1. Summary of model performance across regression, uncertainty-aware, and peak-estimation metrics. All metrics are reported as the median (Q2), with sub- and superscripts denoting the distance to the 25th (Q1) and 75th (Q3) percentiles, respectively. Bold values indicate the best-performing model for each metric. We note that the better performance of BNN on MPIW is in fact, problematic: it indicates the model has high precision, but also low accuracy (poor PICP). Similarly, GPs' performance on forecasting peak time is driven by an inherent bias (see Sec. 3.1).

	Metric	BNN	MLP (pytorch)	MLP (sklearn)	NF	GP	GP2	GP3	ANP
Standard Regression	MSE	$0.46^{+1.36}_{-0.29}$	$0.46^{+1.34}_{-0.28}$	$0.48^{+1.30}_{-0.29}$	$0.60^{+1.98}_{-0.39}$	$0.49^{+1.34}_{-0.30}$	$0.37^{+1.04}_{-0.23}$	$0.37^{+1.07}_{-0.24}$	$\mathbf{0.18}^{+0.50}_{-0.10}$
	MSSE	$7.95^{+11.28}_{-4.28}$	$8.06^{+11.30}_{-4.08}$	$8.48^{+10.71}_{-4.24}$	$10.41^{+17.76}_{-5.64}$	$8.55^{+11.24}_{-4.52}$	$6.60^{+9.16}_{-3.44}$	$6.56^{+9.36}_{-3.46}$	$\mathbf{3.01}^{+3.53}_{-1.38}$
	MAPE	748^{+774}_{-272}	784^{+758}_{-279}	729^{+660}_{-276}	1126^{+1316}_{-435}	1089^{+1208}_{-470}	888^{+968}_{-374}	878^{+963}_{-371}	$\mathbf{294}^{+220}_{-82}$
	MASE	$2.21^{+1.32}_{-0.64}$	$2.25^{+1.26}_{-0.60}$	$2.27^{+1.23}_{-0.59}$	$2.55^{+1.63}_{-0.75}$	$2.42^{+1.61}_{-0.79}$	$2.09^{+1.29}_{-0.63}$	$2.08^{+1.29}_{-0.63}$	$\mathbf{1.37}^{+0.61}_{-0.27}$
Uncertainty-Aware Regression	χ^2 ($\times 10^3$)	787^{+801}_{-368}	210^{+357}_{-114}	205^{+366}_{-111}	$10.9^{+16.9}_{-5.5}$	$2.98^{+1.57}_{-0.87}$	$2.28^{+1.59}_{-0.70}$	$2.32^{+1.60}_{-0.71}$	$\mathbf{1.02}^{+0.34}_{-0.16}$
	$\mathrm{NRMSE_p}$	$31.3^{+11.3}_{-8.4}$	$15.6^{+10.3}_{-4.9}$	$15.4^{+10.2}_{-4.9}$	$3.61^{+2.14}_{-1.06}$	$1.90^{+0.42}_{-0.27}$	$1.68^{+0.44}_{-0.27}$	$1.69^{+0.44}_{-0.27}$	$\mathbf{1.13}^{+0.14}_{-0.07}$
	$\mathrm{NRMSE_o}$	$4.58^{+4.23}_{-1.79}$	$4.61^{+4.15}_{-1.70}$	$4.70^{+4.01}_{-1.68}$	$5.27^{+5.34}_{-2.07}$	$4.68^{+4.09}_{-1.75}$	$4.10^{+3.68}_{-1.53}$	$4.13^{+3.73}_{-1.59}$	$\mathbf{2.82}^{+2.46}_{-0.94}$
	NLPD	496^{+425}_{-230}	123^{+219}_{-66}	120^{+221}_{-66}	$5.84^{+10.42}_{-3.40}$	$1.94^{+1.12}_{-0.76}$	$1.36^{+1.15}_{-0.71}$	$1.40^{+1.17}_{-0.75}$	$\mathbf{0.06}^{+0.45}_{-0.21}$
	CRPS	$0.38^{+0.43}_{-0.15}$	$0.38^{+0.44}_{-0.16}$	$0.39^{+0.45}_{-0.17}$	$0.40^{+0.63}_{-0.18}$	$0.34^{+0.33}_{-0.13}$	$0.29^{+0.31}_{-0.11}$	$0.29^{+0.31}_{-0.11}$	$\mathbf{0.18}^{+0.17}_{-0.05}$
	$\mathrm{PICP}_{0.95}$ (%)	$14.2^{+9.3}_{-4.8}$	$31.0^{+4.5}_{-4.5}$	$31.7^{+4.2}_{-4.0}$	$69.6^{+5.2}_{-7.4}$	$88.3^{+2.2}_{-3.3}$	$87.1^{+2.8}_{-3.6}$	$86.9^{+2.8}_{-3.6}$	$\mathbf{92.7}^{+1.6}_{-2.6}$
	$\mathrm{MPIW}_{0.95}$	$\mathbf{0.10}^{+0.11}_{-0.04}$	$0.22^{+0.06}_{-0.04}$	$0.22^{+0.05}_{-0.04}$	$0.78^{+0.19}_{-0.07}$	$1.81^{+1.43}_{-0.66}$	$1.44^{+1.27}_{-0.46}$	$1.43^{+1.30}_{-0.49}$	$1.01^{+0.47}_{-0.19}$
Peak-Estimation (days / mag)	$\lvert\Delta t_{\mathrm{peak}}\rvert$	27^{+20}_{-12}	28^{+20}_{-13}	27^{+20}_{-12}	47^{+26}_{-25}	13^{+8}_{-5}	12^{+8}_{-4}	13^{+7}_{-5}	$\mathbf{8}^{+6}_{-3}$
	$\lvert\Delta m_{\mathrm{peak}}\rvert$	$0.25^{+0.23}_{-0.11}$	$\mathbf{0.24}^{+0.20}_{-0.11}$	$0.24^{+0.19}_{-0.10}$	$0.38^{+0.30}_{-0.19}$	$0.35^{+0.26}_{-0.17}$	$0.34^{+0.27}_{-0.16}$	$0.34^{+0.27}_{-0.16}$	$0.25^{+0.22}_{-0.10}$
	$\lvert\Delta t_{\mathrm{peak}}\rvert$ (per filter)	$58.3^{+13.5}_{-14.0}$	$59.0^{+14.0}_{-14.0}$	$57.7^{+15.0}_{-15.7}$	$62.7^{+17.0}_{-15.7}$	$45.2^{+14.7}_{-13.2}$	$48.0^{+14.0}_{-15.0}$	$47.4^{+13.4}_{-14.8}$	$\mathbf{18.2}^{+15.5}_{-9.0}$
	$\lvert\Delta m_{\mathrm{peak}}\rvert$ (per filter)	$0.31^{+0.20}_{-0.12}$	$0.30^{+0.18}_{-0.11}$	$0.32^{+0.17}_{-0.11}$	$0.35^{+0.23}_{-0.14}$	$0.34^{+0.19}_{-0.12}$	$0.31^{+0.18}_{-0.10}$	$0.31^{+0.18}_{-0.11}$	$\mathbf{0.23}^{+0.15}_{-0.07}$
	Forecasting $\lvert\Delta t_{\mathrm{peak}}\rvert$	27^{+25}_{-13}	28^{+21}_{-12}	28^{+20}_{-12}	61^{+33}_{-25}	14^{+14}_{-6}	14^{+12}_{-6}	14^{+12}_{-6}	$\mathbf{7}^{+4}_{-2}$
	Forecasting $\lvert\Delta t_{\mathrm{peak}}\rvert$ (per filter)	$73.7^{+16.0}_{-16.0}$	$74.2^{+17.3}_{-16.2}$	$71.7^{+17.7}_{-16.7}$	$89.5^{+29.5}_{-19.2}$	$48.7^{+16.7}_{-15.8}$	$43.7^{+19.7}_{-16.7}$	$44.3^{+20.2}_{-16.7}$	$\mathbf{16.8}^{+15.5}_{-7}$

to learn a prior, it is computationally intensive. But this provides significant speed-up at inference, when a prediction only requires a forward pass, unlike the benchmarks which all must be individually initialized and (re)fit to every single light curve in the test set. Because of this, inference with our ANP on the test set takes $(4.92 \pm 2.38) \times 10^{-6}$ seconds per light curve on two NVIDIA T4 GPUs (evaluated over 80 runs processing parallelized batches of 500 light curves). This is four orders of magnitude faster than the next-best method (MLP; $\sim 5.7 \times 10^{-2}$ s / light curve), and over five orders of magnitude faster than GPs (~ 1.2 s / light curve), the current standard.

At this speed, the ANP is uniquely capable for application to LSST alerts ($\sim$7 million every night), with a single night's prediction computable in just over a minute. This allows for *on-line* use with the LSST - making it possible to generate uncertainty-aware representations that evolve with time, and support time-domain astrophysics for the next ten years.

4. Conclusion & Discussion

We developed an Attentive Neural Process implementation specifically designed for sparse, multifilter astronomical light curves typical of astronomical surveys. We benchmarked it against seven models commonly used in astronomy following a well-established paradigm (Demianenko et al., 2023). We tested all models across 17 metrics, and demonstrated that **the ANP consistently outperforms all tested benchmark models**. Whether measuring raw regression quality (Tab. 3.1), the recovery of astrophysical features (*e.g.*, peak time and magnitude, Tab. 3.1), or rigorous probabilistic calibration (*e.g.*, CRPS, PIT histograms, Sec. 3.2), the ANP excels. Simultaneously, the ANP is highly calibrated, and avoids both the over-conservative uncertainties of GPs and the catastrophic overconfidence of standard neural networks, making it a reliable probabilistic predictor.

Through meta learning and amortized inference, the ANP attains microsecond inference per light curve on a single GPU; >4 orders-of-magnitude faster than benchmark models (Sec. 3.3). Thus the ANP is the first method to clear all three bars — speed, calibration, and class-agnosticism — making it uniquely capable for on-line deployment on the live LSST alert stream. Entirely data-driven, attention-based, and morphologically agnostic across 14 transient classes, it is well-positioned to support time-domain astronomy through the next decade with LSST.

Insights for methods development. A few observations from this work generalize beyond astronomy to other ir-

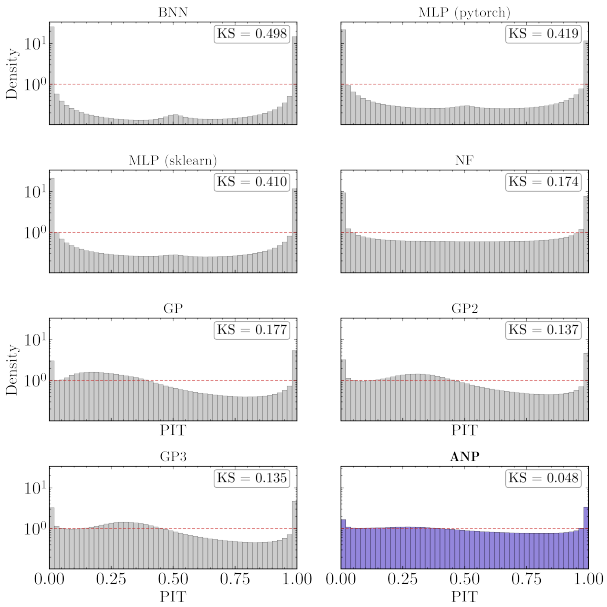

Figure 4. Probability Integral Transform (PIT) histograms for all eight models. For a perfectly calibrated model, PIT values follow Uniform$(0, 1)$ – a flat histogram (red dashed line at density $= 1$). The KS statistic (lower is better; max distance between empirical and uniform CDF) is shown in each panel. ANP (KS=0.048) vastly outperforms the other methods; MLPs and BNN show catastrophic boundary spikes; GPs show a left-skewed bias.

regular, multi-channel scientific time series. First, we implemented a treating discrete channels (filters, sensors) as input *coordinates*, which allows for proper representation of sparse irregular data, allowing relations between channels to be learnt from data alone. Second, we demonstrate that even the simple train-time augmentations we implemented (re-zeroing inputs) can provide translation equivariance in ANPs, allowing them to learn patterns in relative time. Finally, we note that attention-based aggregation is necessary on heterogeneous classes, which motivated our choise of ANP within the NP family; the mean-pooled CNP and NP underfit; attention allows for improved expression when dealing with multiple classes. Our provided software is explicitly designed-agnostic of science-use case, working with meta-datasets consisting of multivariate populations of time series from different classes.

4.1. Limitations

We trained and tested our models on synthetic data derived from the PLAsTiCC (Hložek et al., 2023; PLAsTiCC Modelers, 2022) models. These simulations are widely accepted as highly effective at representing the physical morphologies and noise distributions expected from LSST. The immediate next step is to deploy and evaluate this framework on true observational datasets, such as those from the Zwicky Transient Facility (ZTF) and early LSST alerts.

While our dataset contained different explosive transient classes, the ability of ANP to reconstruct light-curves driven by radically different physical processes remains to be demonstrated: can the same ANP model correctly supernovae and transiting planets?

Because of the ANP's natural ability to learn a robust, continuous representation from irregular multifilter data, its latent space ($\vec{z}$, Sec. 2) encodes a rich, dimensionally reduced summary of the light curve. We are yet to exploit this latent space for inference: we expect this latent representation may serve as an exceptionally powerful feature space for downstream scientific tasks, including physical parameter estimation, similarity searches, and the identification of true novelties ('unknown unknowns') in the imminent LSST data stream to accelerate discoveries in time-domain astrophysics.

While we have justified our choice of ANP (Sec. 2 and Sec. 4), other NP flavours could be implemented for multivariate sparse time series and should be compared with ANP as well as neural ODE methods such as SELDON (Wu et al., 2026), once implemented in a class-agnostic framework.

Our implementation of ANP did not take the observational uncertainty as input. Observational uncertainties are complex but exceptionally well understood in most astronomical surveys and extremely valuable in inference. Relatively small modifications of our architectures would allow the integration of observational uncertainties in the context points and we leave this to future work.

The four astronomical adaptations (filter-as-coordinate, dynamic slicing, temporal re-zeroing, stratified sampling) were chosen for documented engineering reasons (ragged-tensor handling, GPU memory, translation-equivariance, class imbalance) rather than empirical sweeps. We defer component-wise ablations to future work.

Acknowledgements

SC acknowledges support received from the University of Delaware Doctoral Fellowship of Excellence and the NASA FINESST program, Grant 80NSSC25K0312. FBB is supported in part by NSF AST Award Numbers 2511639 and 2308016. AAM is supported in part by NASA Grant 80NSSC24M0020.

SC and FBB are grateful to Andjelka B. Kovačević and Dragana Ilić for insightful conversations on neural processes; Emille Ishida and David Jones, along with the PLAsTiCC Team, for the simulated data. SC is also grateful to Konstantin Malanchev, Oleksandra Razim and members of FASTLab for helpful discussions which helped shape this work.

The authors acknowledge the support of the Vera C. Rubin

Legacy Survey of Space and Time Science Collaborations[6] and particularly of the Transient and Variable Star Science Collaboration[7] (TVS SC) that provided opportunities for collaboration and exchange of ideas and knowledge.

Finally, the authors are grateful to free and open-source software which supports this research, including but not limited to: `SNANA` (Kessler et al., 2009), `fulu` (Demianenko et al., 2023), `rubin_sim` (Yoachim et al., 2026), `LightCurveLynx` (Dai et al., 2026), `Keras` (Chollet et al., 2015), `TensorFlow` (Abadi et al., 2015), `jax` (Bradbury et al., 2018), `PyTorch` (Paszke et al., 2019), `NumPy` (Harris et al., 2020), `Jupyter` (Kluyver et al., 2016), `matplotlib` (Hunter, 2007), `seaborn` (Waskom, 2021), `scikit-learn` (Pedregosa et al., 2011), `SciPy` (Virtanen et al., 2020), `pandas` (Wes McKinney, 2010; pandas development team, 2020), `polars` (Vink et al., 2026), `tqdm` (da Costa-Luis, 2019), and `Python3` (Van Rossum & Drake, 2009).

Impact Statement

This work develops a probabilistic interpolator for astronomical light curves, motivated by the needs of science with Rubin Observatory's LSST. Its impact is concentrated in the time-domain astronomy community: calibrated, ultrafast reconstruction enables online uncertainty aware analysis of the LSST alert stream ($\sim 10^7$ events/night), all of which directly support discovery in transient astrophysics. We see no foreseeable negative societal impact – our model operates on astronomical data, involves no human subjects, and makes no decisions about people. The training cost is modest (5 hours on two NVIDIA T4 GPUs), and we release code and trained weights to support reproducibility. We note that the underlying techniques – attention-based meta-learning over irregular, multi-channel scientific time series – should generalize beyond astronomy to other physical-science domains where sparse multivariate observations are the norm.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

Aigrain, S. and Foreman-Mackey, D. Gaussian process regression for astronomical time series. *Annual Review of Astronomy and Astrophysics*, 61(1):329–371, 2023.

Andersson, T. R., Bruinsma, W. P., Markou, S., Requeima, J., Coca-Castro, A., Vaughan, A., Ellis, A.-L., Lazzara, M. A., Jones, D., Hosking, S., and Turner, R. E. Environmental sensor placement with convolutional Gaussian neural processes. *Environmental Data Science*, 2:e32, 2023. ISSN 2634-4602. doi: 10.1017/eds.2023.22. URL https://www.cambridge.org/core/product/identifier/S2634460223000225/type/journal_article.

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S. S., Arango, S. P., Kapoor, S., Zschiegner, J., Maddix, D. C., Wang, H., Mahoney, M. W., Torkkola, K., Wilson, A. G., Bohlke-Schneider, M., and Wang, Y. Chronos: Learning the Language of Time Series, 2024. URL https://arxiv.org/abs/2403.07815. Version Number: 3.

Bhardwaj, K., Christov, A., and Karpov, S. A Photometric Classifier for Tidal Disruption Events in Rubin LSST, September 2025. URL http://arxiv.org/abs/2509.25902. arXiv:2509.25902 [astro-ph].

Blunt, S., Carvalho, A., David, T. J., Beichman, C., Zink, J. K., Gaidos, E., Behmard, A., Bouma, L. G., Cody, D., Dai, F., Foreman-Mackey, D., Grunblatt, S., Howard, A. W., Kosiarek, M., Knutson, H. A., Rubenzahl, R. A., Beard, C., Chontos, A., Giacalone, S., Hirano, T., Johnson, M. C., Lubin, J., Akana Murphy, J. M., Petigura, E. A., Van Zandt, J., and Weiss, L. Overfitting Affects the Reliability of Radial Velocity Mass Estimates of the V1298 Tau Planets. *The Astronomical Journal*, 166(2):62, August 2023. ISSN 0004-6256, 1538-3881. doi: 10.3847/1538-3881/acde78. URL https://iopscience.iop.org/article/10.3847/1538-3881/acde78.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Katariya, Y., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. JAX: composable transformations of Python+NumPy programs, 2018. URL http://github.com/jax-ml/jax.

Bruinsma, W. P., Requeima, J., Foong, A. Y. K., Gordon, J., and Turner, R. E. The Gaussian Neural Process, January 2021. URL http://arxiv.org/abs/2101.03606. arXiv:2101.03606 [stat].

[6] https://www.lsstcorporation.org/science-collaborations

[7] https://lsst-tvssc.github.io/

Bruinsma, W. P., Markou, S., Requiema, J., Foong, A. Y. K., Andersson, T. R., Vaughan, A., Buonomo, A., Hosking, J. S., and Turner, R. E. Autoregressive Conditional Neural Processes, March 2023. URL http://arxiv.org/abs/2303.14468. arXiv:2303.14468 [stat].

Cabrera-Vives, G., Moreno-Cartagena, D., Astorga, N., Reyes-Jainaga, I., Förster, F., Huijse, P., Arredondo, J., Muñoz Arancibia, A. M., Bayo, A., Catelan, M., Estévez, P. A., Sánchez-Sáez, P., Álvarez, A., Castellanos, P., Gallardo, P., Moya, A., and Rodriguez-Mancini, D. ATAT: Astronomical Transformer for time series and Tabular data. *Astronomy and Astrophysics*, 689:A289, September 2024. ISSN 0004-6361. doi: 10.1051/0004-6361/202449475. URL https://ui.adsabs.harvard.edu/abs/2024A&A...689A.289C.

Chaini, S., Mahabal, A., Kembhavi, A., and Bianco, F. B. Light curve classification with DistClassiPy: A new distance-based classifier. *Astronomy and Computing*, 48:100850, July 2024. ISSN 2213-1337. doi: 10.1016/j.ascom.2024.100850. URL https://ui.adsabs.harvard.edu/abs/2024A&C....4800850C. ADS Bibcode: 2024A&C....4800850C.

Chollet, F. et al. Keras. https://keras.io, 2015.

Clocchiatti, A., Wheeler, J., Phillips, M., Suntzeff, N., Cristiani, S., Phillips, A., Harkness, R., Dopita, M., Beuermann, K., Rosa, M., et al. Sn 1983v in ngc 1365 and the nature of stripped envelope core-collapse supernovae. *The Astrophysical Journal*, 483(2):675–697, 1997.

Cvorovic-Hajdinjak, I. Modeling quasar variability through self-organizing map-based neural process. *Serbian Astronomical Journal*, (208):17–27, 2024. ISSN 1450-698X, 1820-9289. doi: 10.2298/SAJ2408017C. URL https://doiserbia.nb.rs/Article.aspx?ID=1450-698X2408017C.

da Costa-Luis, C. O. 'tqdm': A fast, extensible progress meter for python and cli. *Journal of Open Source Software*, 4(37):1277, 2019. doi: 10.21105/joss.01277. URL https://doi.org/10.21105/joss.01277.

Dai, M., Kubica, J., Malanchev, K., Malz, A. I., Lynn, O., Connolly, A., Mandelbaum, R., and Wood-Vasey, W. M. LightCurveLynx: Forward Modeling of Time-Domain Surveys with Application to ZTF SN Ia DR2, April 2026. URL https://ui.adsabs.harvard.edu/abs/2026arXiv260407134D. ADS Bibcode: 2026arXiv260407134D.

Das, A., Kong, W., Sen, R., and Zhou, Y. A decoder-only foundation model for time-series forecasting, 2023. URL https://arxiv.org/abs/2310.10688. Version Number: 4.

Demianenko, M., Malanchev, K., Samorodova, E., Sysak, M., Shiriaev, A., Derkach, D., and Hushchyn, M. Understanding of the properties of neural network approaches for transient light curve approximations. *Astronomy and Astrophysics*, 677:A16, September 2023. ISSN 0004-6361. doi: 10.1051/0004-6361/202245189. URL https://ui.adsabs.harvard.edu/abs/2023A&A...677A..16D. ADS Bibcode: 2023A&A...677A..16D.

Donoso-Oliva, C., Becker, I., Protopapas, P., Cabrera-Vives, G., Vishnu, M., and Vardhan, H. ASTROMER. A transformer-based embedding for the representation of light curves. *Astronomy and Astrophysics*, 670: A54, February 2023. ISSN 0004-6361. doi: 10.1051/0004-6361/202243928. URL https://ui.adsabs.harvard.edu/abs/2023A&A...670A..54D.

Donoso-Oliva, C., Becker, I., Protopapas, P., Cabrera-Vives, G., Cádiz-Leyton, M., and Moreno-Cartagena, D. Astromer 2, February 2025. URL https://ui.adsabs.harvard.edu/abs/2025arXiv250202717D.

Dubois, Y., Gordon, J., and Foong, A. Y. Neural process family. http://yanndubs.github.io/Neural-Process-Family/, September 2020.

Foreman-Mackey, D., Agol, E., Ambikasaran, S., and Angus, R. Fast and scalable Gaussian process modeling with applications to astronomical time series. *The Astronomical Journal*, 154(6):220, December 2017. ISSN 0004-6256, 1538-3881. doi: 10.3847/1538-3881/aa9332. URL http://arxiv.org/abs/1703.09710. arXiv:1703.09710 [astro-ph].

Garnelo, M., Rosenbaum, D., Maddison, C. J., Ramalho, T., Saxton, D., Shanahan, M., Teh, Y. W., Rezende, D. J., and Eslami, S. M. A. Conditional Neural Processes, July 2018a. URL http://arxiv.org/abs/1807.01613. arXiv:1807.01613 [cs].

Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Eslami, S. M. A., and Teh, Y. W. Neural Processes, July 2018b. URL http://arxiv.org/abs/1807.01622. arXiv:1807.01622 [cs].

Garza, A., Challu, C., and Mergenthaler-Canseco, M. TimeGPT-1, 2023. URL https://arxiv.org/abs/2310.03589. Version Number: 3.

Gebhard, T. D., Angerhausen, D., Konrad, B. S., Alei, E., Quanz, S. P., and Schölkopf, B. Parameterizing pressure–temperature profiles of exoplanet atmospheres with neural networks. *Astronomy & Astrophysics*, 681:A3, January 2024. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/202346390. URL https://www.aanda.org/10.1051/0004-6361/202346390.

Gneiting, T. and Raftery, A. E. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477): 359–378, March 2007. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214506000001437. URL http://www.tandfonline.com/doi/abs/10.1198/016214506000001437.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. Probabilistic Forecasts, Calibration and Sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(2):243–268, April 2007. ISSN 1369-7412, 1467-9868. doi: 10.1111/j.1467-9868.2007.00587.x. URL https://academic.oup.com/jrsssb/article/69/2/243/7109375.

Gordon, J., Bruinsma, W. P., Foong, A. Y. K., Requeima, J., Dubois, Y., and Turner, R. E. Convolutional Conditional Neural Processes, June 2020. URL http://arxiv.org/abs/1910.13556. arXiv:1910.13556 [stat].

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. On calibration of modern neural networks. In *International conference on machine learning*, pp. 1321–1330. PMLR, 2017. doi: 10.48550/arXiv.1706.04599.

Hamad, H. A. and Rosenbaum, D. Flow Matching Neural Processes, December 2025. URL http://arxiv.org/abs/2512.23853. arXiv:2512.23853 [cs].

Hambleton, K. M., Bianco, F. B., Street, R., Bell, K., Buckley, D., Graham, M., Hernitschek, N., Lund, M. B., Mason, E., Pepper, J., Prsa, A., Rabus, M., Raiteri, C. M., Szabo, R., Szkody, P., Andreoni, I., Antoniucci, S., Balmaverde, B., Bellm, E., Bonito, R., Bono, G., Botticella, M. T., Brocato, E., Bricman, K. B., Cappellaro, E., Carnerero, M. I., Chornock, R., Clarke, R., Cowperthwaite, P., Cucchiara, A., D'Ammando, F., Dage, K. C., Dall'Ora, M., Davenport, J. R. A., de Martino, D., de Somma, G., Criscienzo, M. D., Stefano, R. D., Drout, M., Fabrizio, M., Fiorentino, G., Gandhi, P., Garofalo, A., Giannini, T., Gomboc, A., Greggio, L., Hartigan, P., Hundertmark, M., Johnson, E., Johnson, M., Jurkic, T., Khakpash, S., Leccia, S., Li, X., Magurno, D., Malanchev, K., Marconi, M., Margutti, R., Marinoni, S., Mauron, N., Molinaro, R., Moller, A., Moniez, M., Muraveva, T., Musella, I., Ngeow, C.-C., Pastorello, A., Petrecca, V., Piranomonte, S., Ragosta, F., Reguitti, A., Righi, C., Ripepi, V., Sandoval, L. R., Stassun, K. G., Stroh, M., Terreran, G., Trimble, V., Tsapras, Y., van Velzen, S., Venuti, L., and Vink, J. S. Rubin observatory lsst transients and variable stars roadmap, 2022. URL https://arxiv.org/abs/2208.04499.

Hamuy, M. Review on the observed and physical properties of core collapse supernovae. In *2003 Aspen Summer Workshop on the Nuclear Physics of Core Collapse Supernovae*, 1 2003.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

Healy, B. F., Coughlin, M. W., Mahabal, A. A., Laz, T. J. d., Drake, A., Graham, M. J., Hillenbrand, L. A., Roestel, J. v., Szkody, P., Zielske, L., Guiga, M., Hassan, M. Y., Hughes, J. L., Nir, G., Parikh, S., Park, S., Purohit, P., Rebbapragada, U., Reed, D., Warshofsky, D., Wold, A., Bloom, J. S., Masci, F. J., Riddle, R., and Smith, R. The ZTF Source Classification Project: III. A Catalog of Variable Sources, April 2024. URL http://arxiv.org/abs/2312.00143. arXiv:2312.00143 [astro-ph].

Hložek, R., Malz, A. I., Ponder, K. A., Dai, M., Narayan, G., Ishida, E. E. O., Jr, T. A., Bahmanyar, A., Bi, X., Biswas, R., Boone, K., Chen, S., Du, N., Erdem, A., Galbany, L., Garreta, A., Jha, S. W., Jones, D. O., Kessler, R., Lin, M., Liu, J., Lochner, M., Mahabal, A. A., Mandel, K. S., Margolis, P., Martínez-Galarza, J. R., McEwen, J. D., Muthukrishna, D., Nakatsuka, Y., Noumi, T., Oya, T., Peiris, H. V., Peters, C. M., Puget, J. F., Setzer, C. N., Siddhartha, Stefanov, S., Xie, T., Yan, L., Yeh, K.-H., and Zuo, W. Results of the Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC). *The Astrophysical Journal Supplement Series*, 267(2):25, August 2023. ISSN 0067-0049, 1538-4365. doi: 10.3847/1538-4365/accd6a. URL https://iopscience.iop.org/article/10.3847/1538-4365/accd6a.

Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

Hurlbert, S. H. Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs*, 54(2):187–211, June 1984. ISSN 0012-9615, 1557-7015. doi: 10.2307/1942661. URL https://esajournals.onlinelibrary.wiley.com/doi/10.2307/1942661.

Ishida, E. and Jones, D. PERFECT PLASTICC SIM, October 2025. URL https://zenodo.org/doi/10.5281/zenodo.17392460.

Ivezić, Z., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, J. R. P., Angeli, G. Z., Ansari, R., Antilogus, P., Araujo, C., Armstrong, R., Arndt, K. T., Astier, P., Aubourg, E., Auza, N., Axelrod, T. S., Bard, D. J., Barr, J. D., Barrau, A., Bartlett, J. G., Bauer, A. E., Bauman, B. J., Baumont, S., Bechtol, E., Bechtol, K., Becker, A. C., Becla, J., Beldica, C., Bellavia, S., Bianco, F. B., Biswas, R., Blanc, G., Blazek, J., Blandford, R. D., Bloom, J. S., Bogart, J., Bond, T. W., Booth, M. T., Borgland, A. W., Borne, K., Bosch, J. F., Boutigny, D., Brackett, C. A., Bradshaw, A., Brandt, W. N., Brown, M. E., Bullock, J. S., Burchat, P., Burke, D. L., Cagnoli, G., Calabrese, D., Callahan, S., Callen, A. L., Carlin, J. L., Carlson, E. L., Chandrasekharan, S., Charles-Emerson, G., Chesley, S., Cheu, E. C., Chiang, H.-F., Chiang, J., Chirino, C., Chow, D., Ciardi, D. R., Claver, C. F., Cohen-Tanugi, J., Cockrum, J. J., Coles, R., Connolly, A. J., Cook, K. H., Cooray, A., Covey, K. R., Cribbs, C., Cui, W., Cutri, R., Daly, P. N., Daniel, S. F., Daruich, F., Daubard, G., Daues, G., Dawson, W., Delgado, F., Dellapenna, A., de Peyster, R., de Val-Borro, M., Digel, S. W., Doherty, P., Dubois, R., Dubois-Felsmann, G. P., Durech, J., Economou, F., Eifler, T., Eracleous, M., Emmons, B. L., Fausti Neto, A., Ferguson, H., Figueroa, E., Fisher-Levine, M., Focke, W., Foss, M. D., Frank, J., Freemon, M. D., Gangler, E., Gawiser, E., Geary, J. C., Gee, P., Geha, M., Gessner, C. J. B., Gibson, R. R., Gilmore, D. K., Glanzman, T., Glick, W., Goldina, T., Goldstein, D. A., Goodenow, I., Graham, M. L., Gressler, W. J., Gris, P., Guy, L. P., Guyonnet, A., Haller, G., Harris, R., Hascall, P. A., Haupt, J., Hernandez, F., Herrmann, S., Hileman, E., Hoblitt, J., Hodgson, J. A., Hogan, C., Howard, J. D., Huang, D., Huffer, M. E., Ingraham, P., Innes, W. R., Jacoby, S. H., Jain, B., Jammes, F., Jee, M. J., Jenness, T., Jernigan, G., Jevremović, D., Johns, K., Johnson, A. S., Johnson, M. W. G., Jones, R. L., Juramy-Gilles, C., Jurić, M., Kalirai, J. S., Kallivayalil, N. J., Kalmbach, B., Kantor, J. P., Karst, P., Kasliwal, M. M., Kelly, H., Kessler, R., Kinnison, V., Kirkby, D., Knox, L., Kotov, I. V., Krabbendam, V. L., Krughoff, K. S., Kubánek, P., Kuczewski, J., Kulkarni, S., Ku, J., Kurita, N. R., Lage, C. S., Lambert, R., Lange, T., Langton, J. B., Le Guillou, L., Levine, D., Liang, M., Lim, K.-T., Lintott, C. J., Long, K. E., Lopez, M., Lotz, P. J., Lupton, R. H., Lust, N. B., MacArthur, L. A., Mahabal, A., Mandelbaum, R., Markiewicz, T. W., Marsh, D. S., Marshall, P. J., Marshall, S., May, M., McKercher, R., McQueen, M., Meyers, J., Migliore, M., Miller, M., and Mills, D. J. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *The Astrophysical Journal*, 873:111, March 2019. ISSN 0004-637X. doi: 10.3847/1538-4357/ab042c. URL https://ui.adsabs.harvard.edu/abs/2019ApJ...873..111I. ADS Bibcode: 2019ApJ...873..111I.

Jainaga, I. R., Arredondo, J., Valenzuela, C., Rodriguez, D., Sanchez, P., and Heise, P. H. alercebroker/lc_classifier: Release 1.2.3-P, August 2021. URL https://zenodo.org/record/5275453.

Jankov, I., Kovacevic, A. B., Ilić, D., Popović, L. C., Radović, V., Čvorović-Hajdinjak, I., Nikutta, R., and Sánchez-Sáez, P. Photoreverberation mapping of quasars in the context of Legacy Survey of Space and Time observing strategies. *Astronomische Nachrichten*, 343(1-2):e210090, January 2022. ISSN 0004-6337, 1521-3994. doi: 10.1002/asna.20210090. URL https://onlinelibrary.wiley.com/doi/10.1002/asna.20210090.

Kashi, A. and Soker, N. Type ii intermediate-luminosity optical transients (ilots). *Monthly Notices of the Royal Astronomical Society*, 467(3):3299–3305, 2017.

Kasliwal, M. M., Kulkarni, S., Gal-Yam, A., Nugent, P. E., Sullivan, M., Bildsten, L., Yaron, O., Perets, H. B., Arcavi, I., Ben-Ami, S., et al. Calcium-rich gap transients in the remote outskirts of galaxies. *The Astrophysical Journal*, 755(2):161, 2012.

Kessler, R., Bernstein, J. P., Cinabro, D., Dilday, B., Frieman, J. A., Jha, S., Kuhlmann, S., Miknaitis, G., Sako, M., Taylor, M., and Vanderplas, J. SNANA: A Public Software Package for Supernova Analysis. *Publications of the Astronomical Society of the Pacific*, 121:1028, September 2009. ISSN 0004-6280. doi: 10.1086/605984. URL https://ui.adsabs.harvard.edu/abs/2009PASP..121.1028K. ADS Bibcode: 2009PASP..121.1028K.

Khakpash, S., Bianco, F. B., Modjaz, M., Fortino, W. F., Gagliano, A., Larison, C., and Pritchard, T. A. Multifilter UV to Near-infrared Data-driven Light-curve Templates for Stripped-envelope Supernovae. *The Astrophysical Journal Supplement Series*, 275(2):37, December 2024. ISSN 0067-0049, 1538-4365. doi: 10.3847/1538-4365/ad7eaa. URL https://iopscience.iop.org/article/10.3847/1538-4365/ad7eaa.

Kim, H., Mnih, A., Schwarz, J., Garnelo, M., Eslami, A., Rosenbaum, D., Vinyals, O., and Teh, Y. W. Attentive Neural Processes, July 2019. URL http://arxiv.org/abs/1901.05761. arXiv:1901.05761 [cs].

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., and Willing, C. Jupyter notebooks – a publishing format for reproducible computational workflows. In Loizides, F. and Schmidt, B. (eds.), *Positioning and Power in Academic*

Publishing: Players, Agents and Agendas, pp. 87 – 90. IOS Press, 2016.

Kovacevic, A. B., Ilic, D., Popovic, L. C., Mitrovic, N. A., Nikolic, M., Pavlovic, M. S., Hajdinjak, I. C., Knezevic, M., and Savic, D. V. Deep learning of quasar lightcurves in the LSST era, June 2023. URL http://arxiv.org/abs/2306.09357. arXiv:2306.09357 [astro-ph].

Li, W., Chen, H.-Y., Rehemtulla, N., Shah, V. G., Wu, D., Kim, D., Lin, Q., Miller, A. A., and Liu, H. StarEmbed: Benchmarking Time Series Foundation Models on Astronomical Observations of Variable Stars, January 2026. URL http://arxiv.org/abs/2510.06200. arXiv:2510.06200 [astro-ph].

Lupton, R. H., Gunn, J. E., and Szalay, A. S. A Modified Magnitude System that Produces Well-Behaved Magnitudes, Colors, and Errors Even for Low Signal-to-Noise Ratio Measurements. *The Astronomical Journal*, 118 (3):1406–1410, September 1999. ISSN 00046256. doi: 10.1086/301004. URL https://iopscience.iop.org/article/10.1086/301004.

Magill, D., Nicholl, M., Anilkumar, V., van Velzen, S., Sheng, X., Mai, T. S., Tran, H. V., Doan, N. P., Moore, T., Srivastav, S., Young, D. R., Angus, C. R., and Weston, J. MALLORN: many artificial LSST light curves based on observations of real nuclear transients. *RAS Techniques and Instruments*, 5:rzag019, January 2026. ISSN 2752-8200. doi: 10.1093/rasti/rzag019. URL https://academic.oup.com/rasti/article/doi/10.1093/rasti/rzag019/8512818.

Mahabal, A., Sheth, K., Gieseke, F., Pai, A., Djorgovski, S. G., Drake, A. J., and Graham, M. J. Deep-learnt classification of light curves. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–8, Honolulu, HI, November 2017. IEEE. ISBN 978-1-5386-2726-6. doi: 10.1109/SSCI.2017.8280984. URL http://ieeexplore.ieee.org/document/8280984/.

Malanchev, K. light-curve: Light curve analysis toolbox. *Astrophysics Source Code Library*, pp. ascl:2107.001, July 2021. URL https://ui.adsabs.harvard.edu/abs/2021ascl.soft07001M. ADS Bibcode: 2021ascl.soft07001M.

Malanchev, K. L., Pruzhinskaya, M. V., Korolev, V. S., Aleo, P. D., Kornilov, M. V., Ishida, E. E. O., Krushinsky, V. V., Mondon, F., Sreejith, S., Volnova, A. A., Belinski, A. A., Dodin, A. V., Tatarnikov, A. M., Zheltoukhov, S. G., and (The SNAD Team). Anomaly detection in the Zwicky Transient Facility DR3. *Monthly Notices of the Royal Astronomical Society*, 502(4):5147–5175, March 2021. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stab316. URL https://academic.oup.com/mnras/article/502/4/5147/6129794.

Mortimer, P., Diaconu, C., Rochussen, T., Mlodozeniec, B., and Turner, R. E. Incremental Transformer Neural Processes, February 2026. URL http://arxiv.org/abs/2602.18955. arXiv:2602.18955 [cs].

Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., and Hložek, R. RAPID: Early Classification of Explosive Transients Using Deep Learning. *Publications of the Astronomical Society of the Pacific*, 131(1005):118002, November 2019. ISSN 0004-6280, 1538-3873. doi: 10.1088/1538-3873/ab1609. URL https://iopscience.iop.org/article/10.1088/1538-3873/ab1609.

Möller, A. and de Boissière, T. SuperNNova: an open-source framework for Bayesian, neural network-based supernova classification. *Monthly Notices of the Royal Astronomical Society*, 491(3):4277–4293, January 2020. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stz3312. URL https://academic.oup.com/mnras/article/491/3/4277/5651173.

Nguyen, T. and Grover, A. Transformer Neural Processes: Uncertainty-Aware Meta Learning Via Sequence Modeling, February 2023. URL http://arxiv.org/abs/2207.04179. arXiv:2207.04179 [cs].

pandas development team, T. pandas-dev/pandas: Pandas, February 2020. URL https://doi.org/10.5281/zenodo.3509134.

Park, J. W., Villar, A., Li, Y., Jiang, Y.-F., Ho, S., Lin, J. Y.-Y., Marshall, P. J., and Roodman, A. Inferring Black Hole Properties from Astronomical Multivariate Time Series with Bayesian Attentive Neural Processes, June 2021. URL http://arxiv.org/abs/2106.01450. arXiv:2106.01450 [astro-ph].

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. *PyTorch: an imperative style, high-performance deep learning library*. Curran Associates Inc., Red Hook, NY, USA, 2019.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Pellegrino, C., Pritchard, T. A., Modjaz, M., Crawford, A., Khakpash, S., and Bianco, F. GOPREAUX I: Open-source Code and Data to Model Multi-wavelength Emission of Extragalactic Transients using Gaussian Processes. *arXiv e-prints*, art. arXiv:2604.03372, April 2026. doi: 10.48550/arXiv.2604.03372.

Peng, Y., Ristić, M., Kedia, A., O'Shaughnessy, R., Fontes, C. J., Fryer, C. L., Korobkin, O., Mumpower, M. R., Villar, V. A., and Wollaeger, R. T. Kilonova Light-Curve Interpolation with Neural Networks. *Physical Review Research*, 6(3):033078, July 2024. ISSN 2643-1564. doi: 10.1103/PhysRevResearch.6.033078. URL http://arxiv.org/abs/2402.05871. arXiv:2402.05871 [astro-ph].

Pion, A. and Vazquez, E. Design-marginal calibration of Gaussian process predictive distributions: Bayesian and conformal approaches, December 2025. URL https://centralesupelec.hal.science/hal-05400282.

PLAsTiCC Modelers. Libraries & Recommended Citations for using PLAsTiCC Models, June 2022. URL https://zenodo.org/record/2612895.

Qu, H. and Sako, M. Photometric Classification of Early-time Supernova Light Curves with SCONE. *The Astronomical Journal*, 163:57, February 2022. ISSN 0004-6256. doi: 10.3847/1538-3881/ac39a1. URL https://ui.adsabs.harvard.edu/abs/2022AJ....163...57Q. ADS Bibcode: 2022AJ....163...57Q.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass., 3. print edition, 2008. ISBN 978-0-262-18253-9.

Rubin's Survey Cadence Optimization Committee, Bianco, F. B., Jones, R. L., Anguita, T., Bauer, F. E., Edwards, L. O. V., Jha, S. W., Mandelbaum, R., Miller, A. A., Olsen, K., Slater, C. T., Smartt, S. J., Strader, J., Street, R. A., Volk, K., and Yoachim, P. Survey Cadence Optimization Committee's Phase 3 Recommendations. Project Science Technical Note PSTN-056, NSF-DOE Vera C. Rubin Observatory, January 2025. URL https://pstn-056.lsst.io/.

Ruiter, A. J. Type ia supernova sub-classes and progenitor origin. *Proceedings of the International Astronomical Union*, 15(S357):1–15, 2019.

Russeil, E., Malanchev, K. L., Aleo, P. D., Ishida, E. E. O., Pruzhinskaya, M. V., Gangler, E., Lavrukhina, A. D., Volnova, A. A., Voloshina, A., Semenikhin, T., Sreejith, S., Kornilov, M. V., and Korolev, V. S. RAINBOW: A colorful approach to multipassband light-curve estimation. *Astronomy & Astrophysics*, 683:A251, March 2024. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/202348158. URL https://www.aanda.org/10.1051/0004-6361/202348158.

Sasseville, G., Hlavacek-Larrondo, J., Haggard, D., Adam, A., Paugnat, H., and Witzel, G. Probabilistic Interpolation of Sagittarius A*'s Multi-Wavelength Light Curves Using Diffusion Models, January 2026. URL http://arxiv.org/abs/2601.20863. arXiv:2601.20863 [astro-ph].

Scholz, J., Andersson, T. R., Vaughan, A., Requeima, J., and Turner, R. E. Sim2Real for Environmental Neural Processes, October 2023. URL http://arxiv.org/abs/2310.19932. arXiv:2310.19932 [cs].

Schuetz, A.-K., Migala, A., Boesky, A., Poon, A. W. P., Broekgaarden, F. S., and Li, A. RESOLVE: Rare Event Surrogate Likelihood for Gravitational Wave Paleontology Parameter Estimation, June 2025. URL http://arxiv.org/abs/2506.00757. arXiv:2506.00757 [astro-ph].

Shah, V. G., Gagliano, A., Malanchev, K., Narayan, G., Malz, A. I., and and the LSST Dark Energy Science Collaboration. ORACLE: A Real-time, Hierarchical, Deep Learning Photometric Classifier for the LSST. *The Astrophysical Journal*, 995(1):4, December 2025. ISSN 0004-637X, 1538-4357. doi: 10.3847/1538-4357/ae1130. URL https://iopscience.iop.org/article/10.3847/1538-4357/ae1130.

Shen, Y. and Gagliano, A. Diffusion Autoencoders with Perceivers for Long, Irregular and Multimodal Astronomical Sequences, October 2025. URL http://arxiv.org/abs/2510.20595. arXiv:2510.20595 [stat].

Sánchez-Sáez, P., Reyes, I., Valenzuela, C., Förster, F., Eyheramendy, S., Elorrieta, F., Bauer, F. E., Cabrera-Vives, G., Estévez, P. A., Catelan, M., Pignata, G., Huijse, P., De Cicco, D., Arévalo, P., Carrasco-Davis, R., Abril, J., Kurtev, R., Borissova, J., Arredondo, J., Castillo-Navarrete, E., Rodriguez, D., Ruz-Mieres, D., Moya, A., Sabatini-Gacitúa, L., Sepúlveda-Cobo, C., and Camacho-Iñiguez, E. Alert Classification for the ALeRCE Broker System: The Light Curve Classifier. *The Astronomical Journal*, 161(3):141, March 2021. ISSN 0004-6256, 1538-3881. doi: 10.3847/1538-3881/abd5c1. URL https://iopscience.iop.org/article/10.3847/1538-3881/abd5c1.

Van Roestel, J., Duev, D. A., Mahabal, A. A., Coughlin, M. W., Mróz, P., Burdge, K., Drake, A., Graham, M. J., Hillenbrand, L., Bellm, E. C., Kupfer, T., Delacroix, A., Fremling, C., Golkhou, V. Z., Hale,

D., Laher, R. R., Masci, F. J., Riddle, R., Rosnet, P., Rusholme, B., Smith, R., Soumagnac, M. T., Walters, R., Prince, T. A., and Kulkarni, S. R. The ZTF Source Classification Project. I. Methods and Infrastructure. *The Astronomical Journal*, 161(6):267, June 2021. ISSN 0004-6256, 1538-3881. doi: 10.3847/1538-3881/abe853. URL https://iopscience.iop.org/article/10.3847/1538-3881/abe853.

Van Rossum, G. and Drake, F. L. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009. ISBN 1441412697.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention Is All You Need, June 2017. URL http://arxiv.org/abs/1706.03762. arXiv:1706.03762 [cs].

Villar, V. A., Cranmer, M., Berger, E., Contardo, G., Ho, S., Hosseinzadeh, G., and Lin, J. Y.-Y. A Deep-learning Approach for Live Anomaly Detection of Extragalactic Transients. *The Astrophysical Journal Supplement Series*, 255:24, August 2021. ISSN 0067-0049. doi: 10.3847/1538-4365/ac0893. URL https://ui.adsabs.harvard.edu/abs/2021ApJS..255...24V. ADS Bibcode: 2021ApJS..255...24V.

Vink, R., Beedie, A., nameexhaustion, Peters, O., Burghoorn, G., de Gooijer, S., Gorelli, M. E., reswqa, Marshall, van Zundert, J., Denecker, K., Hulselmans, G., Grinstead, C., Sprenkels, A., Patyk, K., Manley, L., chielP, Valtar, K., Turner-Trauring, I., Mitchell, L., Afroozeh, A., Genockey, K., Harbeck, H., eitsupi, Bergdoll, L., Robin, Santamaria, M., Borchert, O., and deanm0000. pola-rs/polars: Python Polars 1.41.2, May 2026. URL https://zenodo.org/doi/10.5281/zenodo.7697217.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Wang, X., Zhao, H., and Bonilla, E. V. Rényi Neural Processes, June 2025. URL http://arxiv.org/abs/2405.15991. arXiv:2405.15991 [cs].

Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021. doi: 10.21105/joss.03021. URL https://doi.org/10.21105/joss.03021.

Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman (eds.), *Proceedings of the 9th Python in Science Conference*, pp. 56 – 61, 2010. doi: 10.25080/Majora-92bf1922-00a.

Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., and Sahoo, D. Unified Training of Universal Time Series Forecasting Transformers, 2024. URL https://arxiv.org/abs/2402.02592. Version Number: 2.

Wu, J., O'Brien, J., Li, J., Krafczyk, M. S., Shah, V. G., Wasserman, A. R., Apley, D. W., Narayan, G., and Samia, N. I. SELDON: Supernova Explosions Learned by Deep ODE Networks, March 2026. URL http://arxiv.org/abs/2603.04392. arXiv:2603.04392 [astro-ph].

Yalavarthi, V. K., Burchert, J., and Schmidt-thieme, L. Tripletformer for Probabilistic Interpolation of Irregularly sampled Time Series. 2022. doi: 10.48550/ARXIV.2210.02091. URL https://arxiv.org/abs/2210.02091. Version Number: 2.

Yoachim, P., Jones, L., Eric H. Neilsen, J., Tiago, Leistedt, B., Parejko, J., Bellm, E., Street, R., Carlin, J., Humna, Becker, M. R., pgris, erykoff, reneehlozek, Prisinzano, L., Gollotti, G. A., Dennihy, E., Abrams, N., lmptc, Sick, J., LI, Assef, R. J., Ross, Bricman, K., Bregeon, J., Lim, K.-T., Guy, L., Kelley, M., and Andreoni, I. lsst/rubin_sim: v2.6.1, January 2026. URL https://zenodo.org/doi/10.5281/zenodo.7087822.

Čvorović-Hajdinjak, I., Kovacevic, A. B., Ilić, D., Popović, L. C., Dai, X., Jankov, I., Radović, V., Sánchez-Sáez, P., and Nikutta, R. Conditional Neural Process for non-parametric modeling of active galactic nuclei light curves. *Astronomische Nachrichten*, 343(1-2):e210103, January 2022. ISSN 0004-6337, 1521-3994. doi: 10.1002/asna.20210103. URL https://onlinelibrary.wiley.com/doi/10.1002/asna.20210103.

A. Dataset

A.1. Astrophysical Simulations

The astrophysical models included in our dataset are:

1. three thermonuclear supernova classes: SN Ia, Ia91bg, Iax. See (Ruiter, 2019) for a review of thermonuclear supernovae including SN Ia and subclasses;
2. four core collapse supernova classes: SN II, IIP, IIL, IIn. A description of the phenomenology and underlying physics of core collapse supernovae may be found in (Hamuy, 2003);
3. five stripped envelope supernova subclasses: SN IIb, Ib, Ic, and Ic-BL. See, *e.g.*, (Clocchiatti et al., 1997);
4. and two other transients: Calcium Rich Transients (CART; Kasliwal et al., 2012) and Intermediate Luminosity Transients (ILOTs; Kashi & Soker, 2017)

The original full dataset is highly imbalanced, with over 300,000 light curves for type II-plateau supernovae (SN IIPs), but only 1,589 intermediate luminosity transients (ILOTs). While our pipeline supports imbalanced datasets (via stratified sampling, Sec. 2.1), we experimented with using hundreds of thousands of light curves and and found that convergence took orders of magnitude longer without any drastic improvement in performance. So, we instead adopt a class-balanced subsample. Full details of our final dataset are given in Tab. 2.

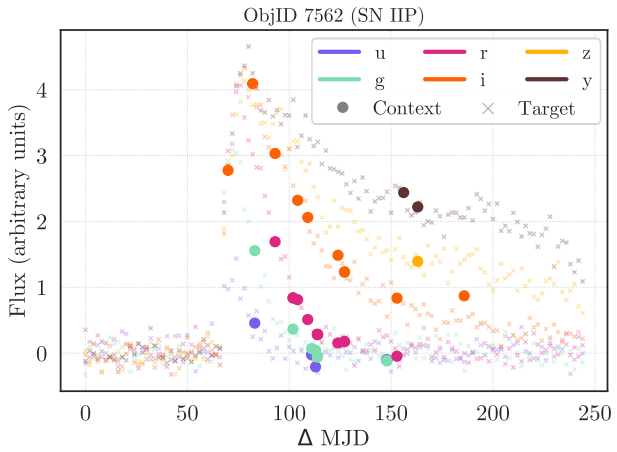

Figure 5. An example of a Type IIP Supernova from our synthetic dataset in the 6 LSST filters ($ugrizy$). The full set of simulated points are denoted by crosses (target points), while a realistic subset that would be observed by LSST is denoted by the circles (context points). Note that this is just a single realization of an LSST cadence based on LSST OpSims v5.1.1. The task of light curve interpolation is to reconstruct the complete multifilter light curve (target points/crosses) from only the sparse observations (context points/circles).

A.2. Time Series Sampling Strategy

To produce LSST-like light curves for our test set, we use the official LSST Operations Simulator (OpSims; Yoachim et al., 2026) version 5.1.1, which simulates the currently planned survey strategy of LSST (Rubin's Survey Cadence Optimization Committee et al., 2025). We place each astrophysical source at a random location (RA, Dec) within the LSST main survey footprint (Wide Fast Deep, (Ivezić et al., 2019)). While the OpSims tables include many features (*e.g.*, observing conditions, camera parameters), we use them only to impose an LSST-like cadence on top of our dense-simulated light curves; we do not consider factors like galactic latitude. The OpSims table gives us the complete lifetime of LSST observations at the given location; we choose the start time of our transient to be a random value within the survey lifetime, and extract the dates corresponding to the OpSims observations where our simulated transient would be active. We linearly interpolate the dense-cadence light curves in each filter to retrieve the LSST mock observations. We repeat this procedure ten times per test light curve to take statistical aggregates across realizations; each realization produces a light curve with between 10 and 141 context points. An example of one LSST-like realization for a light curve is given by the circles in Fig. 5.

B. ANP Implementation

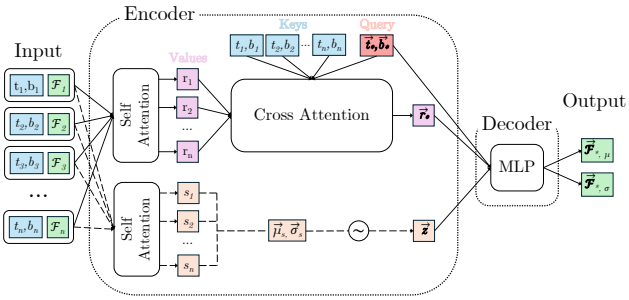

Figure 6. Model architecture of the Attentive Neural Process, adapted from (Kim et al., 2019). The input consists of pairs of values $(\vec{t}_*, \vec{b}_*)$ representing time and wavelength filter, and corresponding flux values $\vec{\mathcal{F}}_*$.

Architecture. A schematic figure representing an ANP has been plotted in Fig. 6. In addition to Sec. 2.1, a complete implementation of our architecture is available in `https://github.com/sidchaini/NightLANP`.

Time complexity. For a single light curve with n_C context points and n_* query points, an ANP forward pass scales as $\mathcal{O}(n_C^2 + n_C n_*)$ (Kim et al., 2019), dominated by self-attention over the context and cross-attention from queries to context. In our setting, $n_C \in [10, 141]$ and n_* is typically a few hundred, so attention computation is negligible com-

Table 2. The dataset used in this work, which is a subsampled version of the simulated light curves in (Ishida & Jones, 2025), based on the PLAsTiCC models (PLAsTiCC Modelers, 2022). Light curves are sampled every 2 days in the six LSST filters (20% of the data is sampled every 10 days).

Transient Physics	Class	PLAsTiCC Model Name	Full Dataset Size	Train Size	Test Size
Thermonuclear	SN Ia	`SALT2.WFIRST-H17`	109,324	1,000	250
	SN Ia-91bg	`SIMSED.SNIa-91bg`	11,979	1,000	250
	SN Iax	`SIMSED.SNIax`	31,620	1,000	250
Stripped-Envelope	SN Ib	`NON1ASED.V19_CC+HostXT`	222,578	1,000	250
	SN Ibc	`SIMSED.SNIbc-MOSFIT`	48,508	1,000	250
	SN Ic	`NON1ASED.V19_CC+HostXT`	183,501	1,000	250
	SN Ic-BL	`NON1ASED.V19_CC+HostXT`	36,088	1,000	250
	SN IIb	`NON1ASED.V19_CC+HostXT`	80,902	1,000	250
Core-Collapse	SN II	`SIMSED.SNII-NMF`	189,800	1,000	250
	SN IIL	`NON1ASED.V19_CC+HostXT`	184,072	1,000	250
	SN IIP	`NON1ASED.V19_CC+HostXT`	302,829	1,000	250
	SN IIn	`NON1ASED.V19_CC+HostXT`	51,753	1,000	250
		`SIMSED.SNIIn-MOSFIT`	14,632	1,000	250
Other	CART	`SIMSED.CART-MOSFIT`	10,777	1,000	250
	ILOT	`SIMSED.ILOT-MOSFIT`	1,589	1,000	250

pared to MLP feature processing. Crucially, this cost is *paid only once*, at training time for the encoder; at inference time, the same trained model is amortized across all light curves – there is no re-fitting, in contrast to GPs which scale as $\mathcal{O}(N^3)$ per object due to kernel matrix inversion. We thus find that the ANP inference takes $(4.0 \pm 0.9) \times 10^{-6}$ seconds per light curve, and is orders of magnitudes faster as compared to the next fastest MLP (sklearn)'s $(5.5 \pm 0.3) \times 10^{-2}$ seconds, and the fastest GP's (1.1 ± 0.2) seconds. All neural benchmarks (BNN, NF, MLP-PyTorch) and the ANP were run on the same pair of NVIDIA T4 GPUs; GP fits and MLP-sklearn run on CPU. The benchmark models were implemented per `fulu` defaults. The empirical speedup observed for ANPs ($\sim 10^4$–$10^5\times$) in Fig. 7 reflects this elimination of per-object fitting rather than hardware differences.

C. More Details on Evaluation Metrics

In Tab. 1, we looked at the median behavior of the ANP against the benchmark models. Here, we look more in depth at the distribution. First, Fig. 8 visualizes the win-rate percentage of the different models for each metric – for all test light curves, what percentage of model 'wins' as the best performer. We see that the ANP is the best-performing by most metrics.

Additionally, all the metrics reported in Tab. 1 have also been visualized here as boxplots to compare the distributions of our metrics for the different models. Boxes span the IQR, and whiskers extend to 1.5×IQR. Outliers omitted for clarity.

- Standard Regression Metrics are plotted in Fig. 9.
- Uncertainty-Aware Regression Metrics are plotted in Fig. 10, 14, 13.
- Peak-Estimation Metrics are plotted in Fig. 11.
- Forecasting Peak-Estimation Metrics are plotted in Fig. 12.

Additionally, the MSSE as a function of broad physical class is also plotted in Fig. 15.

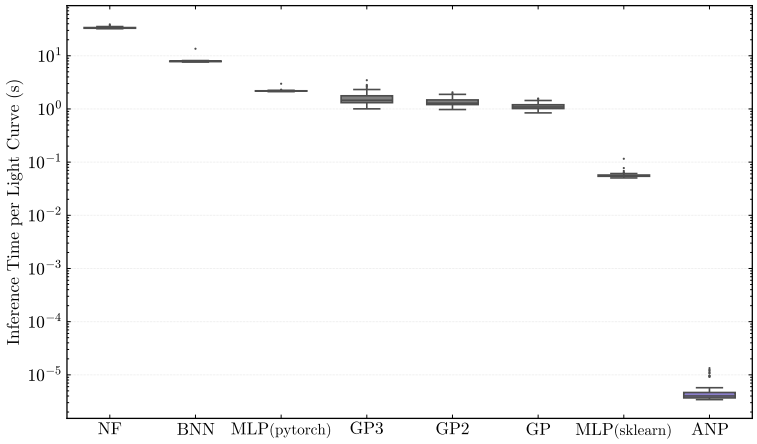

Figure 7. Distribution of per-object inference times (in seconds, log y-axis) for ANP and all seven benchmark models considered in this work (Sec. 3), excluding training time (lower is better). ANP achieves $\sim 10^{-6}$ s/object – four orders of magnitude faster than the next-best neural benchmark (MLP) and over five orders of magnitude faster than GPs.

	BNN	MLP (pytorch)	MLP (sklearn)	NF	GP	GP2	GP3	**ANP**
MSSE	2%	1%	1%	1%	2%	2%	3%	88%
MASE	2%	1%	1%	0%	3%	3%	3%	86%
MAPE	2%	2%	3%	0%	5%	2%	2%	84%
MSE	2%	1%	1%	1%	3%	3%	4%	85%
χ^2	0%	0%	0%	0%	1%	4%	4%	92%
$\mathrm{NRMSE_p}$	0%	0%	0%	0%	1%	4%	4%	92%
$\mathrm{NRMSE_o}$	2%	1%	1%	1%	3%	2%	3%	87%
CRPS	0%	0%	0%	1%	3%	3%	3%	89%
NLPD	0%	0%	0%	0%	1%	2%	3%	94%
$\lvert\Delta t_{\mathrm{peak}}\rvert$ (per filter)	1%	1%	1%	1%	4%	3%	2%	88%
$\lvert\Delta m_{\mathrm{peak}}\rvert$ (per filter)	8%	10%	12%	8%	4%	5%	5%	49%
$\lvert\Delta t_{\mathrm{peak}}\rvert$	5%	4%	5%	2%	16%	10%	2%	55%
$\lvert\Delta m_{\mathrm{peak}}\rvert$	10%	20%	25%	11%	4%	3%	3%	25%
Forecasting $\lvert\Delta t_{\mathrm{peak}}\rvert$	6%	5%	4%	1%	14%	11%	2%	57%
Forecasting $\lvert\Delta t_{\mathrm{peak}}\rvert$ (per filter)	1%	1%	1%	0%	3%	10%	4%	79%

Win rate (%): 0, 20, 40, 60, 80, 100

Figure 8. Per-light-curve win-rate summary across the 13 metrics in Table 1 (higher is better). Each cell reports the fraction of light curves on which that model achieves the lowest error for that metric. ANP wins overwhelmingly on all reconstruction metrics (top 9 rows) and on the four peak-time/magnitude metrics. Only on global peak-magnitude prediction do the MLP variants compete, and ANP remains at least as good as the next-best model.

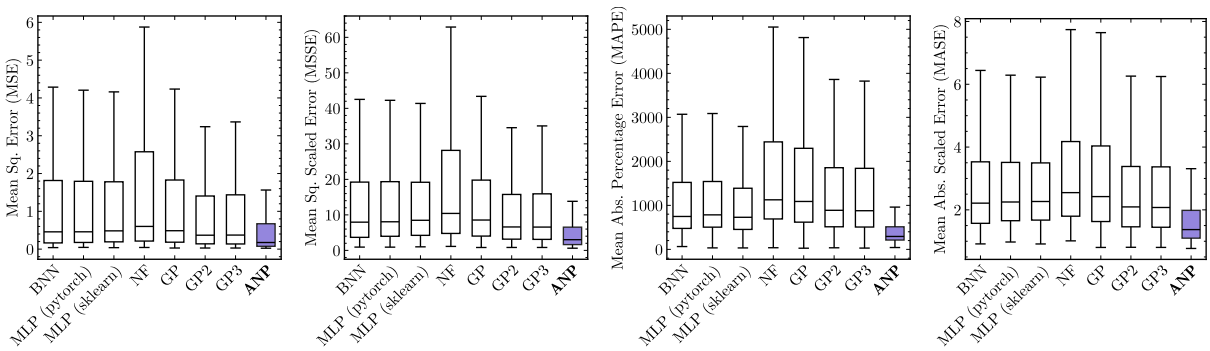

Figure 9. Distribution of four regression-quality metrics (Tab. 3.1) for the unseen 3,750 test light curves (14 transient classes, 250 objects each) across all eight reconstruction models (lower is better). Each light curve includes context points for 10 LSST-cadence realisations. (a) MSE is computed per object across all filters, while (b) MSSE, (c) MAPE, and (d) MASE are computed per object per filter and averaged across filters. All metrics are median-aggregated over the 10 realisations to mitigate pseudoreplication (Hurlbert, 1984). Boxes span the IQR; whiskers extend to 1.5×IQR; outliers omitted for clarity. The ANP (purple, filled) outperforms all models on all four metrics.

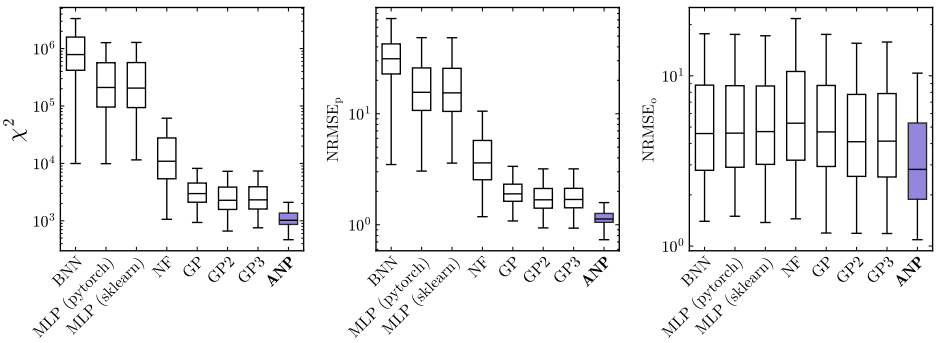

Figure 10. As Fig. 9 for three uncertainty-aware regression metrics (Tab. 3.1); y-axis log scale. ANP (purple, filled) achieves the lowest median and most compact distribution on all three metrics. The ANP distribution entirely outperforms BNN and MLPs on χ^2 and $\mathrm{NRMSE_p}$.

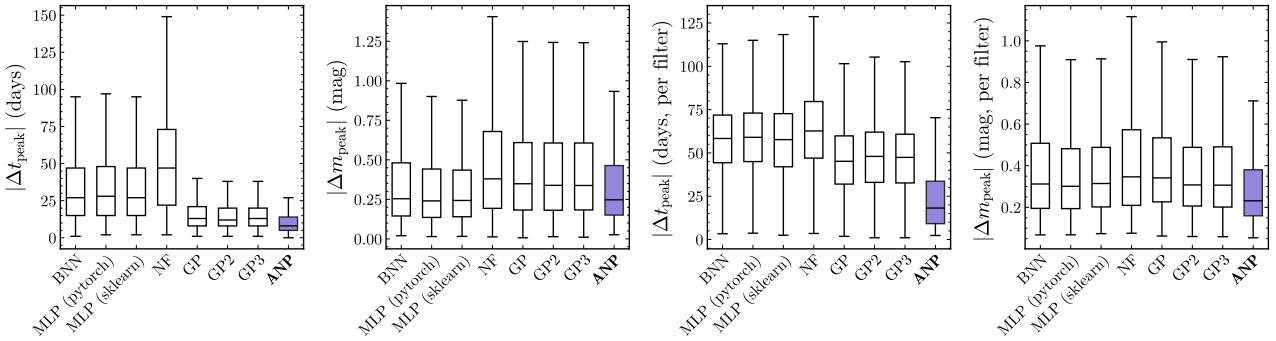

Figure 11. Distribution of four peak-estimation metrics (Tab. 3.1; lower is better). All metrics median-aggregated over 10 realisations. (a) Global peak-time error $|\Delta t_{\mathrm{peak}}|$ in days. (b) Global peak-magnitude error $|\Delta m_{\mathrm{peak}}|$ in luptitudes. (c) Per-filter peak-time error, averaged across filters. (d) Per-filter peak-magnitude error. ANP (purple, filled) is consistently the top performer.

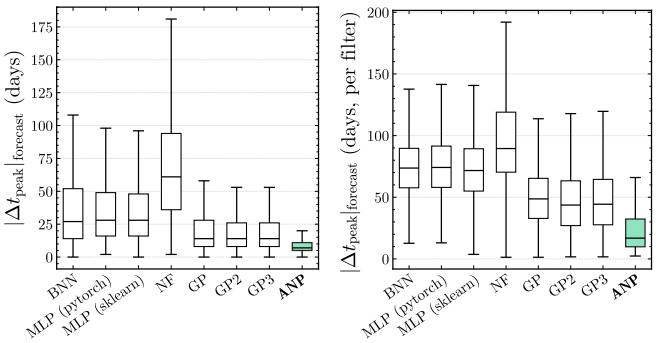

Figure 12. Forecasting peak time from pre-peak context (Tab. 3.1). Left: global peak-time absolute error. Right: per-filter peak-time absolute error averaged across filters. The dedicated forecasting-trained ANP outperforms all other models on both global and per-filter peak forecasting.

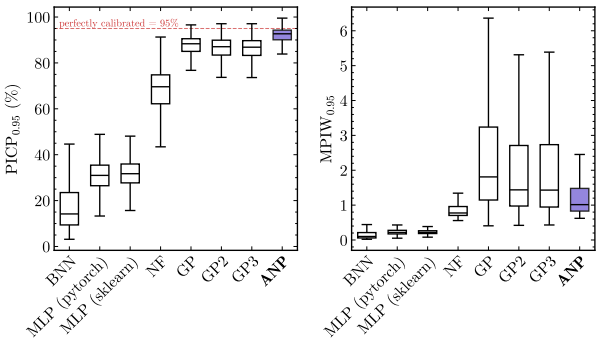

Figure 13. Distribution of two probabilistic evaluation metrics (Sec. 3.2). All metrics median-aggregated over 10 realisations. (a) Prediction Interval Coverage Probability (PICP) at 95% confidence – the fraction of true flux values falling within the 95th-percentile predicted interval. (b) Mean Prediction Interval Width (MPIW) at 95%. Read together, the other neural models (BNN, MLPs, NF) overestimate confidence relative to ANP (high precision, low accuracy).

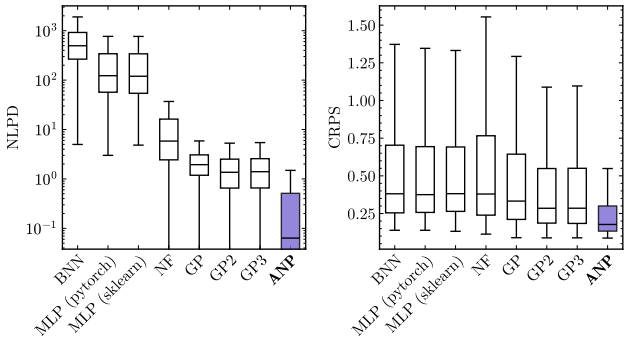

Figure 14. Distribution of two strictly-proper-scoring-rule metrics, following the (Gneiting et al., 2007) framework (lower is better). All metrics median-aggregated over 10 realisations. (a) Negative Log Predictive Density (NLPD), log scale. (b) Continuous Ranked Probability Score (CRPS), a generalisation of MAE to probabilistic forecasts.

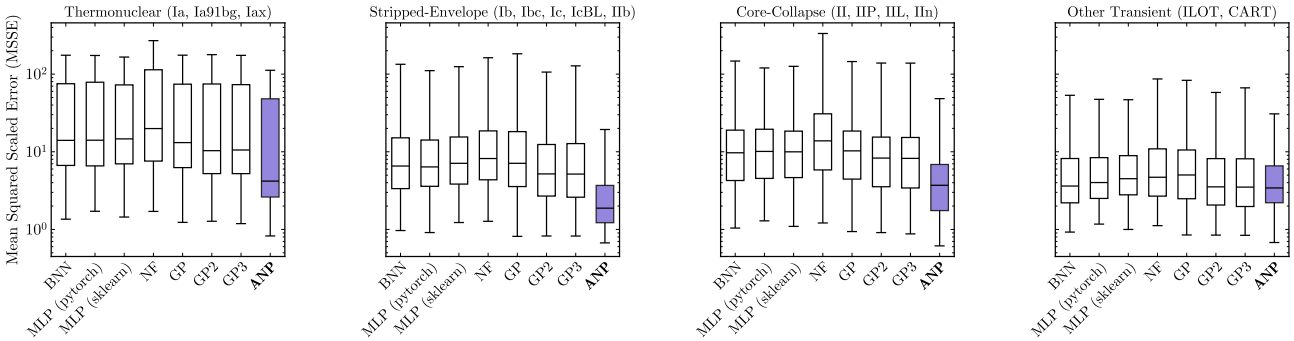

Figure 15. Distribution of MSSE as a function of broad physical class. SN-like transients (Core-Collapse, Stripped-Envelope) are modelled most effectively by ANP. For thermonuclear SNe (more homogeneous), ANP still outperforms but with a smaller margin. For Other transients (ILOT, CART) the ANP performance is statistically comparable to other models.